

ANALYSIS OF STUDENTS' STATISTICAL REASONING IN SOLVING NON-ROUTINE PROBLEMS

Dinda Amalia Rahmah

Mathematics Education, Faculty of Mathematics and Natural Science, Universitas Negeri Surabaya

e-mail: dindarahmah@mhs.unesa.ac.id

Rini Setianingsih

Mathematics Education, Faculty of Mathematics and Natural Science, Universitas Negeri Surabaya

e-mail: rinisetianingsih@unesa.ac.id

Abstract

Statistical reasoning is needed by students to understanding statistical concepts and problems. Students' efforts in understanding statistical information and problems often involve statistical process, namely the process of organizing and reducing data and the process of analyzing and interpreting data. This research is a descriptive study with a qualitative approach that aims to analyze and describe the senior high school students' statistical reasoning in solving non-routine problems. The subjects in this study were three high school students of grade XII who have same gender and have equivalent mathematical abilities. The instruments used in this study consisted of the Mathematical Ability Test (MAT), the Statistical Reasoning Test (SRT), and interview guidelines. The results showed that students who have low mathematical ability are also reached low level in statistical reasoning. In solving problem related to process of organizing and reducing data, the students reached level 1 and level 2 in statistical reasoning, because the students determine the mode value using incorrect concepts and ways, that are input the data to the mode formula incorrectly, determining the mode value of data which are presented in the frequency distribution table just by looking at frequency column, and cannot combine and sort the data in percent form correctly. In solving the problem related to the process of analyzing and interpreting data, the students reached level 1 in statistical reasoning, because they cannot interpret the tendency of two data groups, so the response which is given is incorrect.

Keywords: statistics, statistical process, statistical reasoning.

INTRODUCTION

The reasoning is one aspect that is emphasized in mathematics learning. Reasoning ability is one of determining factors for students' success in learning mathematics. Permendikbud Nomor 59 Tahun 2014 in the appendix III explained that one of mathematics learning goals is to make students can do reasoning, because reasoning ability is needed by students to solve the problems in mathematical context and daily life. Moore and Bruder (1996) added that reasoning ability is needed by students to be able to think clearly and logically. These explanations show that reasoning plays an important role in helping students to think and act logically when solving the problems, both problems related to mathematics and outside of mathematics.

As important as reasoning, statistics today has taken an important role in decision making. Statistics have been considered as guidelines in data collection, data processing, data presentation, as well as in decision making. This statement is in accordance with the definition of statistics, which is as a set of procedures that is used to collect data, analyze data, and make conclusions from data analysis

(Hakim & Kumudji, 1997). Based on these descriptions, it can be said that statistics can be used as a tool to solve problems in daily life, both related to work or the development of science (Moore, 1997).

In line with the important role of statistics in daily life, in Indonesia statistics is taught as part of mathematics at all levels of education. According to Qian (2011), nowadays, understanding concepts and statistical reasoning have become a focus in statistical learning. Statistical reasoning is refers to process of thinking using data and statistical concept, therefore students are able to understand and explain the statistical process, as well as interpret the results of the statistical process (Ben-Zvi & Garfield, 2004).

The results of Rosidah et al. (2018) study related to students' statistical reasoning stated that students who have high and low mathematical ability had good understanding in calculating algorithm using median and standard deviation formula, but shows a low understanding related to the concepts of both, so they cannot interpret the calculation results obtained correctly. These results support the previous study which also revealed that in the analysis and interpretation stage, the male and female students cannot interpret the statistical value obtained according to

the context (Rosidah, 2016). The results above denote that aspect of students' statistical reasoning needs to be further studied and developed.

In this study, statistical reasoning is observed from the results of students answer related to non-routine problem solving that includes the statistical process, specifically the process of organizing and reducing data and the process of analyzing and interpreting data. According to Mooney (2002), the process of organizing and reducing data includes the activity of grouping, organizing, and combine data into simpler forms and also summarizing data using measures of central tendency. According to Langrall and Mooney (2002), the process of analyzing and interpreting data includes the process of identifying trends data to make conclusions and predictions from the data set. Meanwhile, non-routine problems are given because it requires critical thinking and application of some strategies and concepts (Inoue, 2005), so the researcher can assess the ways of students' thinking when applying the statistical and other concepts to solve the problems.

Students' statistical reasoning influenced their ability in applying statistical concepts to solve the problems. In this study, the researcher used statistical reasoning model of Garfield (2002) to classified the level of students' statistical reasoning based on their ability in understanding and applying statistical concepts, namely Level 1 (Idiosyncratic Reasoning), level 2 (Verbal Reasoning), level 3 (Transitional Reasoning), level 4 (Procedural Reasoning), and level 5 (Integrated Reasoning).

The explanation above brings to the question about how the different students' statistical reasoning influence their ability in solving non-routine problems. Therefore, this study was conducted to analyze students' statistical reasoning in solving non-routine problems.

METHODS

This study is descriptive in nature using qualitative approach. It aims to analyze and describe senior high school students' statistical reasoning in solving non-routine problems. This study was conducted in SMA Negeri 1 Balen Bojonegoro in the second semester of the 2018/2019 academic year. The subjects in this study were three students of grade XII natural science program who have equivalent mathematical abilities and same gender.

To collect data, the researchers use Mathematical Ability Tests (MAT), Statistical Reasoning Tests (SRT), and interview guidelines with the following data collection steps: First, the researchers gave MAT to students of grade XII natural science program to select the subjects of study. Based on the MAT scores, three students were selected who have equivalent mathematical abilities and same gender. Second, the researcher gave SRT to obtain the data of students' statistical reasoning in solving non-routine

problems from the subjects of study. In the last, the researchers conducted interview with three subjects of study after the SRT was finished.

After data from SRT and interview were obtained, the researchers analyzed data in three steps, that is data condensation, data presentation, and concluding. The next, the researchers analyzed the students' statistical reasoning in solving non-routine problems based on statistical reasoning indicators in each statistical process which is adapted from Chan et al. (2016).

RESULTS AND DISCUSSION

Data collection in this study was begun by giving Mathematical Ability Tests (MAT) on Wednesday, 27th March 2019 in class XII MIA 2 and on Thursday, 28th March 2019 in class XII MIA 1 of SMA Negeri 1 Balen Bojonegoro. Based on MAT scores, three female students who have equivalent mathematical abilities (all subjects have low mathematical ability) were selected to be subjects of study, namely SLA as first subject (S1), AN as second subject (S2), and DQNF as third subject (S3). After the subjects of study are determined, on Thursday, 25th April 2019 was given Statistical Reasoning Tests (SRT) and interviews to three subjects of study to obtain data about students' statistical reasoning in solving non-routine problems.

To analyze students' statistical reasoning data, the researchers used statistical reasoning model of Garfield (2002) and statistical reasoning indicators adapted from Chan et al. (2016) for the process of organizing and reducing data and the process of analyzing and interpreting data.

Table 1. Indicators of Statistical Reasoning in Problems-solving

Level of Statistical Reasoning	Statistical Process	
	The Process of Organizing and Reducing Data	The Process of Analyzing and Interpreting Data
Level 1 (Idiosyncratic Reasoning)	Organize, combine, or reduce data using the concept of mode value, but completely incorrect	Compare data and make conclusions based on data provided, but completely incorrect
Level 2 (Verbal Reasoning)	Organize, combine, or reduce data using the concept of mode value in words but incomplete or only partially accurate	Compare data and make conclusions based on data provided in words, but incomplete or only partially accurate
Level 3 (Transitional Reasoning)	Organize, combine, or reduce data using the concept of mode value correctly, but unable to relate it to the actual data	Compare data and make conclusions based on data provided correctly, but unable to relate it to the actual data

Level 4 (Procedural Reasoning)	Organize, combine, or reduce data using the concept of mode value correctly, but can't provide full explanation	Compare data and make conclusions based on data provided correctly, but can't provide full explanation
Level 5 (Integrated Reasoning)	Organize, combine, or reduce data using the concept of mode value correctly and able to explain and relate it to the actual data	Compare data and make conclusions based on data provided correctly and able to explain and relate it to the actual data

Students' Statistical Reasoning in the Process of Organizing and Reducing Data

The researchers presented data in Great Britain about age intervals of car and motorcycle drivers, the number of car and motorcycle drivers involved accidents and injured in 2017, as well as the percentage change of the number car and motorcycle drivers involved accidents from 2016 to 2017. Then, the students must determine the age of motorcycle drivers involved the most accidents in the Great Britain in 2016. It aims to assess students' ability in organizing and combining data to obtain new data, and then use them to determine the ageing mode of motorcycle drivers involved accidents in Great Britain in 2016.

Related to the solution of the problem given, S1 and S2 provide the answers as shown below.

Figure 1. S1' answer relating to the process of organizing and reducing data

Figure 2. S2' answer relating to the process of organizing and reducing data

On the statistical reasoning tests' answer sheet (Figure 1 and Figure 2), S1 and S2 do not explain how they determine the age of motorcycle drivers involved the most accidents in the Great Britain in 2016. Then, the researchers conducted interviews with S1 and S2 to find out how they determine that age.

Table 2. Researchers' Interview Transcript with S1 and S2 Related to the Process of Organizing and Reducing Data

S1' Interview Transcript	S2' Interview Transcript
R: What is the first step you do to determine the age of motorcycle drivers involved the most accidents in Great Britain in 2016?	R: What is the first step you do to determine the age of motorcycle drivers involved the most accidents in Great Britain in 2016?
S: (Be quiet, don't answer immediately), directly Miss.	S: I didn't write Miss.

R: How do you determine it?	R: Then, how can you answer like that?
S: Because at this age the number is highest Miss.	S: Because the age of 30-39 years old, the decreased percentage is the lowest.
R: That's the highest number in 2017, do you think in 2016 will be same?	
S: No.	
R: No? Then, how do you determine?	
S: I don't know Miss.	

On the interview transcripts above, it is known that S1 and S2 do not apply mode values concept to determine the age of motorcycle drivers that involved the most accidents in Great Britain in 2016. S1 explains that the reason is motorcycles drivers who involved accidents at 20-29 years old have the highest number. When the researcher confirm her answer about these number is highest of 2017, S1 cannot give the explanation again. From this explanation, it can be known that S1 determine the age of motorcycle drivers who involved the most accidents just by looking at the "frequency" column. Meanwhile, S2 states that the age of 30-39 years old is the age of motorcycle drivers who involved the most accidents in 2016. S2 explains that at the drivers' age of 30-39 years old, the decreased percentage of the number of motorcycle drivers involved accidents from 2016 to 2017 is the lowest. So, S2 considers that at these drivers' age, the number of motorcycle drivers who involved accidents in 2016 is the highest.

S1 and S2' answers show that they cannot combine two data to find new data with suitable concept, namely the concept of fractional percent form, and they do not use the appropriate concept and ways to determine the age of motorcycle drivers who involved the most accidents in 2016. Based on these results, it can be said that S1 and S2 reach statistical reasoning of level 1 in solving non-routine problem that related to the process of organizing and reducing data, because they organize, combine, and determine events which occur the most in data using a concept, but completely incorrect.

Different answers are presented by S3. S3 use mode value concept and calculation about the fractional percent form appropriate with the given data. On the answer sheet and interview which was conducted by the researcher, S3 explains that the first step she calculates the age mode of motorcycle drivers using the formula of mode value. Second, she multiplies the age mode which was obtained from first step with the decreased percentage of the drivers who at 20-29 years old that involved accidents from 2016 to 2017. The last, she subtracts the result of the first step and second step.

Although the concept which is used by S3 is correct (using the formula of mode values), but she cannot

calculates the mode value correctly. She input the incorrect data to the variable of d_2 . She also use the incorrect data from car vehicle and the steps which are used still incorrect. S3' answer sheet is shown as the following picture.

$$Mo = L_0 + \frac{d_1}{d_1 + d_2} \cdot C$$

$$= 19.5 + \frac{5.986}{5.986 + 3.595} \cdot 10$$

$$= 19.5 + \frac{5.986}{9.581} \cdot 10$$

$$= 19.5 + 6.20$$

$$= 25.70$$

$$25.70 \times \frac{8}{100} = 2.4$$

$$= 25 - 2.4$$

$$= 22.6$$

Jadi usia pengemudi sepeda motor yg terlibat kecelakaan terbanyak di Britania Jaya pada tahun 2017 kurang 22,6 tahun.

Figure 3. S3' answer relating to the process of organizing and reducing data

Based on the S3' answers, the researchers can place S3 at the level 2 of statistical reasoning, because she organize, combine, and reduce data using the concept of mode value and other concept in written answer and words, but only partially correct.

Statistical Reasoning in the Process of Analyzing and Interpreting Data

Still use the same data, the students were asked to interpret the tendency of two data groups, and then give the "agree" or "disagree" response about the given statement. One of the student' answer is shown in the table below.

Table 3. One of Student' Answer Relating to the Process of Analyzing and Interpreting Data

S2' Answer	
<p>4. Saya setuju dengan pernyataan kecelakaan mobil lebih beresiko membuat pengemudi terluka daripada kecelakaan sepeda motor karena berdasarkan data diatas di atas jumlah pengemudi yg terluka lebih banyak daripada sepeda motor.</p>	
<p>4. I agree with the statement that the car accidents is more risk to make the drivers injured than the motorcycle accidents because based on the data above, the number of car drivers who injured was greater than motorcycle drivers</p>	

The same response is given by two other students. They said that the car accidents is more risk to make the drivers injured than the motorcycle accidents with the reason that based on data which is listed in the table, the number of car drivers who injured in accidents was greater than the number of motorcycle drivers who injured in accidents.

To determine the risk level of injury due to an accident is by look at the comparison between the number of drivers who injured and the number of drivers who involved the accident. So, the response which is given by students is not appropriate, because they just compare the quantity of the number car drivers and motorcycle drivers who injured, they do not pay attention to the data trends from the comparison of the number of drivers who injured with the drivers who involved in accidents for each vehicle. If this is done, the comparison results of motorcycle will be greater than the car, which means that the motorcycle accidents is more risk to make the drivers injured than car accidents. Therefore, the researchers can place S1, S2, and S3 at the level 1 of statistical reasoning, because they compare data and give response to make conclusions based on data trends, but completely incorrect.

Mathematical Ability Tests (MAT)' scores give the information that three students who were research subjects have low mathematical abilities. This MAT' results in line with the low level of statistical reasoning which are reached by students from Statistical Reasoning Tests (SRT) and interview. National Research Council (2001) explained that there are several mathematical proficiency which are needed by students to study in statistics, namely conceptual understanding (proficiency to integrate some mathematical ideas, for example is ideas of measure which is used in statistical concepts, like the measure of central tendency and measure of spread) , procedural fluency (proficiency to describing data accurately), strategic competence (proficiency to formulate, represent, and solve the statistical related problems), and adaptive reasoning (proficiency to make conclusion based on data trend and situation).

From the explanation above, it can be said that there are relationship between mathematical abilities and statistical abilities, which further can influence the students' statistical reasoning. This statement is supported by Ben-Zvi and Garfield (2004) who said that statistical reasoning is one of statistical abilities' part. Therefore, in the beginning of statistical learning the teachers need to pay attention with the students' abilities in understanding mathematical concepts that underlie statistical concepts.

CLOSURE

Conclusions

Based on the results and data analysis that have been done, it can be concluded that students who have low mathematical ability is also reached low level in statistical reasoning. In the process of organizing and reducing data, the students reached level 1 and 2, because they cannot combine and organize two data to find new data with correct concept, namely the concept of fractional percent form and then use it to determine the mode value. Besides

that, the student use the mode value concept to determine the value that occur the most in data, but input the data to the formula incorrectly (Level 2). The other students determine the value that occur the most in data which are presented in the frequency distribution table just by looking at the "frequency" column and sort the data in percent form incorrectly (Level 1). In the process of analyzing and interpreting data, the students reached level 1, because they just compare the quantity of two data sets and cannot understand the data trends from each data sets, so they cannot interpret and give correct response to the given data.

Suggestions

Based on the study that has been conducted by the researchers, the suggestions which are given as follows.

1. From the results of this study that shows that the statistical reasoning that is done by senior high school students in solving non-routine problems that related to the process of organizing and reducing data and the process of analyzing and interpreting data are still at low level, so is expected that statistical learning in the school can give more attention to understanding statistical concepts and development of statistical reasoning through providing non-routine problems that related to these statistical process.
2. In this study, the researchers analyze the statistical reasoning from the students who have equivalent characteristics, so is expected for the next study to analyze the statistical reasoning from the students who have different characteristics. Therefore, it can be obtained the new knowledge about statistical reasoning students in solving non-routine problems.
3. For the next study that related to students' statistical reasoning, is expected can consideration the weakness of this study and do the next better, so it can give more benefits in statistical learning.

REFERENCES

- Ben-Zvi, D. and Gafield, J. 2004. *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking*. Boston MA: Kluwer Academic Publisher.
- Chan, S.W., Ismail, Z., and Sumintono, B. 2016. "A Framework for Assessing High School Students' Statistical Reasoning". *PLoS ONE*. Vol. 11(11). DOI: 10.1371/journal.pone.0163846.
- Garfield, J.B. 2002. "The Chalange of Developing Statistical Reasoning". *Journal of Statistics Education*, (Online), Vol. 10(3), www.amsat.Org/publicatins/jse/v103/garfield.html, accessed on May 10, 2018).
- Hakim, A. and Kumudji, S. 1997. *Pengantar Statistika*. Surabaya: Citra Media.
- Inoue, N. 2005. The Realistic Reasons Behind Unrealistic Solution: The Role of Interpretive Activity in Word Problem Solving. *Learning and Instruction*, Vol. 15: pp. 69-83.
- Langrall, C. and Mooney, E.S. 2002. "The Development of a Framework Charactizing Middle School Student' Statistical Thinking". *Journal Illinois State University USA*.
- Mooney, E.S. 2002. "A framework for characterizing middle school students' statistical thinking". *Math Think Learn*. Vol. 4: pp. 23-63. DOI: 10.1207/S1 532 7833MTL0401_2.
- Moore, B. and Bruder, K. 1996. *Philosophy: The Power of Ideas (3rd Ed.)*. U. S. A.: Mayfield Publishing.
- Moore, D.S. 1997. "New Pedagogy and New Content: The Case of Statistics". *International Statistics Review*. Vol. 65(2): pp. 123-165.
- Permendikbud No. 59 Tahun 2014 Tentang Kurikulum 2013 Sekolah Menengah Atas/ Madrasah Aliyah. Jakarta: Kemendikbud.
- Qian, G. 2011. "Teaching, Learning and Retention of Statistical Ideas in Introductory Statistics Education". *European Journal of Pure and Applied Mathematics*. Vol. 4(2): pp. 103-116.
- Rosidah. 2016. "Penalaran Statistis siswa SMA dalam Pemecahan Masalah Statistika ditinjau dari Perbedaan gender". *Prosiding Seminar Nasional*. Vol. 2 (1): pp. 57-65.
- Rosidah, Budayasa, I.K., and Juniati, D. 2018. "An Analysis of Statistical Reasoning Process of High School Students in Solving the Statistical Problem". *Journal of Physics: Conference Series*.